

AD-A053 841

AUTONETICS DOWNEY CALIF

F/G 6/4

LEARNING AND EVALUATION OF PATTERN CHARACTERISTICS FOR AUTOMATI--ETC(U)

1962 G G RICKER

EM-1162-141

NL

UNCLASSIFIED

[OF]

AD  
A053841



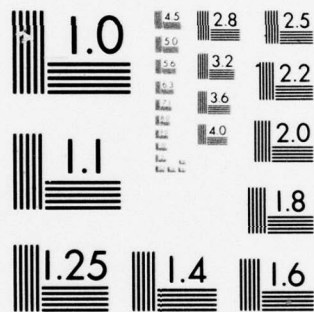
END

DATE

FILMED

6 -78

DDC



MICROCOPY RESOLUTION TEST CHART  
NATIONAL BUREAU OF STANDARDS-1963-A

AD A 053841

AD No. —  
DDC FILE COPY

MOST Project 3

14/ EIA-1162-141  
COPY

2

AUTONETICS  
TECHNICAL LIBRARY

11/ 1962

12/ 46p.

6 LEARNING AND EVALUATION  
OF PATTERN CHARACTERISTICS  
FOR AUTOMATIC RECOGNITION.

By

10/ G.G. RICKER

1962

APPROVED BY:

*R. M. Ashby*  
R. M. ASHBY

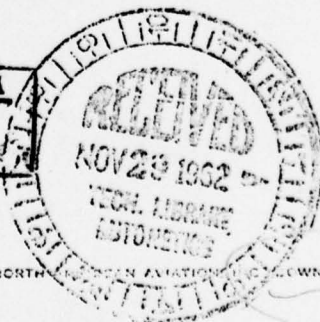
VICE PRESIDENT

9/ RESEARCH AND DEVELOPMENT rept,



DISTRIBUTION STATEMENT A

Approved for public release  
Distribution Unlimited



DDC  
RECEIVED  
APR 28 1978

13 Autonetics

A DIVISION OF NORTH AVIATION CORP., WHEAT RIDGE, CALIFORNIA

048 154

Ym

CHANGE	ADD	PURGE	INPUT WORKSHEET
ACC <u>EM-1162-141</u>	DELETE	REMOVE	
ISS	MF	DLIM	CAC
INDAT	GBSL	CAT	
PROC	PDAT	CLAS	DGRD
TYPE	SNAM		
SRCE1	DOC1	BRCL	
SRCE2	DOC2		
SRCE3	DOC3		
TITL	LEARNING AND EVALUATION OF PATTERN CHARACTERISTICS FOR AUTOMATIC RECOGNITION		
STITL			
AUTH			
CONT			
ACQ			
AVAIL			
NOTE			
MISC			
LOC			
DESC			
IDEN			
ABST			

EM 1162-141

CONTENTS

<u>Section</u>	<u>Page</u>
Illustrations . . . . .	iv
Tables . . . . .	v
Introduction . . . . .	1
I. STATISTICAL DECISIONS . . . . .	3
1. Decision Rule With Cost Independent of Decision . . . . .	3
2. Decision Rule With Cost Dependent Upon Decision . . . . .	5
II. STATISTICAL MODEL . . . . .	ii
1. Model Without Learning . . . . .	11
2. Model With Learning . . . . .	13
3. Information Theory . . . . .	20
4. An Experiment With Written Numerals . . . . .	22
III. CONCLUSIONS . . . . .	39
References . . . . .	40

ACCESSION No.	
NTM	White Section <input checked="" type="checkbox"/>
DOC	Buff Section <input type="checkbox"/>
UNCLASSIFIED	<input type="checkbox"/>
JUSTIFICATION	
<i>Hutter on file</i>	
BY	
EXTENSION/AVAILABILITY CODES	
ONE, TWO, OR SPECIAL	
A	

## ILLUSTRATIONS

<u>Figure</u>		<u>Page</u>
1.	Printed Numerals Used in Experiment . . . . .	23
2.	The Number 4 Quantitized Into a 16 x 16 Cellular Matrix. This is the Number 4 Which Appears in the Last Column of the Hand Printed Numerals . . . .	24
3.	A Compilation of the Values for $\sum_{k=1}^k X_{ik}$ for the Hand Printed 4s Occuring in the First Six Columns in Figure 1. (Specifically These are Values for $\sum_{k=1}^6 X_{4k}$ ) . . . . .	26
4.	Transformed Pattern for the Vertical Scan Lines of the Number 4 Shown in Figure 2 . . . . .	27
5.	Transformed Pattern for the Horizontal Scan Lines of the Number 4 Shown in Figure 2 . . . . .	27
6.	$\sum_{k=1}^k X_{ik}$ for the Vertical Scan Pattern of the 4s . . . . .	27
7.	$\sum_{k=1}^k X_{ik}$ for the Horizontal Scan Pattern of the 4s . . . . .	27
8.	Transinformation Values as Computed from Equation (10) for Each of the Cells in the 16 x 16 Matrix . . . . .	32
9.	The Blackened Cells Have a Transinformation Value Equal to or Greater Than 0.30 (See Figure 8) . . . . .	33
10.	The Blackened Cells Have a Transinformation Equal to or Less Than 0.13 (See Figure 8) . . . . .	34

EM 1162-141

# TABLES

<u>Number</u>		<u>Page</u>
1.	Recognition Matrix for the Seventh Set of Patterns Formed by Summing the Log of the Probabilities for the 64 Cells of the Horizontal Transformation	
	$\sum_{k=1}^{64} \log p_k \left[ y(c)/w_i \right] \dots\dots\dots$	28
2.	Recognition Matrix for the Seventh Set of Patterns Formed by Summing the Log of the Probabilities for the 32 Cells of the Vertical Transformation	
	$\sum_{k=1}^{32} \log p_k \left[ y(c)/w_i \right] \dots\dots\dots$	29
3.	Recognition Matrix for the Combined Results of Tables 1 and 2	30
4.	Recognition Matrix for the Seventh Set of Patterns Formed by Summing the Probabilities for the 256 Cells as Developed from the Learning Set of Patterns	
	$\sum_{k=1}^{256} p_k \left[ y(c)w_i \right] \dots\dots\dots$	35
5.	Recognition Matrix for the Seventh Set of Patterns Formed by Summing the Probabilities for the Blackened 55 Cells of Figure 9	
	$\sum_{55 \text{ cells}} p_k \left[ y(c)/w_i \right] \dots\dots\dots$	36
6.	Recognition Matrix for the Seventh Set of Patterns Formed by Summing the Probability for the Blackened Cells Shown in Figure 10	
	$\sum_{55 \text{ cells}} p_k \left[ y(c)/w_i \right] \dots\dots\dots$	37

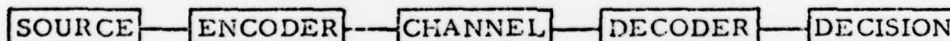
EM 1162-141

## INTRODUCTION

The design of an automatic pattern recognition system involves considerations in three distinct but not totally independent problem areas. The broad selection of pattern characteristics and the design of equipment which will measure these characteristics quickly and efficiently are two aspects of the system which must be resolved. The third is the formulation of a recognition scheme which will establish the significant characteristics and operate upon them to form reliable pattern classifications. This report is concerned primarily with the classification problem and discusses some applications of "statistical decision theory" and "information theory" to the design of a recognition scheme. However, one section of this report presents the results of an experiment and the other problem areas <sup>which</sup> were considered to the extent that characteristics had to be defined and measured before the data could be collected.

In general, the two theories mentioned above complement each other. The decision theory provides the method for optimizing the rules of classification to be applied to the information in the data, and the information theory provides the method for optimizing the rate with which information is presented to the decision rule.

The general structure of a recognition system would appear thus:



The source is the origin of the pattern. The encoder performs an operation upon the pattern which makes it compatible with some specified channel of transmission. The decoder reverses the process of the encoder and can be considered as the element which measures the pattern characteristics. The decision element then classifies the pattern. Recognition is based upon the measurement of a set of characteristics and patterns with similar characteristics categorized in one class. Learning is employed by observing and recording these characteristics for a finite number of patterns from each class.

A classical example of a recognition system is the man-to-man communication through written symbols. A man originates the pattern and codes it in written form, then another man decodes the pattern and classifies it. Of course, noise usually occurs somewhere in the system prior to decoding and complicates the decision process. Too much noise produces an illegible pattern.

1 BEST AVAILABLE COPY

EM 1162-141

The rules of decision discussed below are generally referred to as Bayes' rules in which the objective is to perform the classification in a manner which will minimize the penalty of an incorrect decision. Two penalties or loss functions are considered. In the first case the losses which will be suffered by each decision are established prior to making a decision; in the second case the loss is a function of the decision.

It is generally desirable to develop the probability distributions of the decision rule through a mathematical model of the pattern. The model chosen for this discussion is representative of a binary pattern or signal with multiplicative noise. The probability distributions are developed for both the condition of no learning and the condition in which learning observations are used to enhance the knowledge of the statistics of the model. The case in which learning is not applied to the model follows closely the development by Braverman (Reference 1).

The decision rules presented here evaluate the importance of each pattern or signal characteristic only on a comparative basis between any two classes. In order to evaluate the importance of each characteristic with reference to all classes i.e., the relative importance within the concepts of a recognition system, it is proposed below that serious consideration be given to the expected "mutual-information" of each characteristic as it is defined in information theory.

**BEST AVAILABLE COPY**

## I. STATISTICAL DECISIONS

## 1. DECISION RULE WITH COST INDEPENDENT OF DECISION

The signal or pattern space will be designated as

$\Omega = [w_1, w_2, \dots, w_n]$  where  $w_i$  represents the  $i^{\text{th}}$  class of signals or patterns. The probability distribution over  $\Omega$  will be

$P = [p(w_1), p(w_2), \dots, p(w_n)]$ . Let  $X_i = [X_i(1), X_i(2), \dots, X_i(v)]$  be the vector valued random variable of the characteristics of the transmitted signal  $X_i$  where  $X_i(c)$  is the value of the  $c^{\text{th}}$  characteristic and  $X_i \in w_i$ . The value associated with a random variable will be denoted by the lower case. Let  $Y = [Y(1), Y(2), \dots, Y(v)]$  be the set for the received signal and let the set of decisions associated with a received signal,  $Y$ , be  $D = [d_1, d_2, \dots, d_n]$ . The cost function  $W(w_j/w_i)$  is a measure of the loss incurred with a decision which classifies a signal from the  $i^{\text{th}}$  class as a signal from the  $j^{\text{th}}$  class. The risk,  $R$ , involved in making a decision from  $D$  then depends upon the cost function and the probability distribution for  $D$  given a received signal  $y$ .

$$R[w_i, D(y)] = \sum_{j=1}^n W(w_j/w_i) p(d_j/y)$$

The expected risk depends upon the probability distribution for  $y$  and  $w_i$ .

$$E_R = \sum_Y \sum_{i=1}^n \sum_{j=1}^n W(w_j/w_i) p(d_j/y) p(y, w_i)$$

or,

$$E_R = \sum_Y \sum_{i=1}^n \sum_{j=1}^n W(w_j/w_i) p(d_j/y) p(y/w_i) p(w_i)$$

The object of a decision rule is to minimize this expected risk. A decision which will minimize the risk for each received signal,  $y$ , will also possess a minimized risk when summed over all  $y$ . Thus, it is sufficient to choose a decision rule which will minimize the following term:

EM 1162-141

$$\sum_{i=1}^n \sum_{j=1}^n W(w_j/w_i) p(d_j/y) p(y/w_i) p(w_i).$$

If it is assumed that the decision for a rejection can be ignored; i.e., only recognition or misrecognition are considered valid decisions, and if in addition a misrecognition is assumed equally detrimental for all classes, then,

$$W(w_j/w_i) = 0 \text{ for } i=j$$

$$W(w_j/w_i) = 1 \text{ for } i \neq j$$

With these values for the cost function the risk becomes:

$$R [w_i, D(y)] = \sum_{\substack{j=1 \\ j \neq i}}^n p(d_j/y) = \text{probability of misrecognition.}$$

The expected risk then is.

$$E_R = \sum_{i=1}^n p(y/w_i) p(w_i) \left[ \sum_{\substack{j=1 \\ j \neq i}}^n p(d_j/y) \right].$$

Expanding over the subscript i gives,

$$\begin{aligned} E_R &= p(y/w_1) p(w_1) \sum_{j=2}^n p(d_j/y) \\ &+ p(y/w_2) p(w_2) \sum_{\substack{j=1 \\ j \neq 2}}^n p(d_j/y) \\ &\dots \dots \dots \\ &+ p(y/w_n) p(w_n) \sum_{j=1}^{n-1} p(d_j/y). \end{aligned}$$

In order to minimize the expected risk a decision rule must be chosen which will eliminate the largest component in the above equation. This rule can be formulated as follows:

DECISION: Choose class  $k$  with probability 1 and all other classes with probability 0 such that

$$p(d_k/y) = 1$$

and

$$p(d_j/y) = 0 \text{ for all } j \neq k$$

$$\text{when } p(w_k)p(y/w_k) \geq p(w_j)p(y/w_j) \quad (1)$$

for all  $j \neq k$ .

This is generally referred to as a Bayes' decision rule and quite often written in the form,

$$\frac{p(y/w_k)}{p(y/w_j)} \geq \frac{p(w_j)}{p(w_k)} \quad (2)$$

## 2. DECISION RULE WITH COST DEPENDENT UPON DECISION

Another expression for the cost function is,

$$-\log p(w_i/d_j).$$

The expected risk is then,

$$E_R = - \sum_Y \sum_{i=1}^n \sum_{j=1}^n \log p(w_i/d_j) p(d_j/y) p(y/w_i) p(w_i).$$

However, the following probability expression can be rearranged as

$$\begin{aligned} p(d_j/y) p(y/w_i) p(w_i) &= p(d_j/y) p(w_i, y) = p(d_j/y) p(w_i/y) p(y) \\ &= p(d_j, w_i/y) p(y) = p(d_j, w_i, y) \end{aligned}$$

Therefore,

$$\sum_Y p(d_j/y)p(y/w_i)p(w_i) = \sum_Y p(d_j, w_i, y) = p(d_j, w_i) = p(d_j/w_i)p(w_i)$$

Thus the expected risk can be rewritten as:

$$E_R = - \sum_{i=1}^n \sum_{j=1}^n p(w_i)p(d_j/w_i) \log p(w_i/d_j)$$

This is precisely the definition of equivocation entropy in "information theory." With this cost function, a minimization of the expected risk involves a minimization of "equivocation." In order to compute the conditions for an extremum, when any pair of signals are being considered, divide the observation space into two fields,  $\lambda_1$  and  $\lambda_2$ , then choose  $d_1$  if  $y \in \lambda_1$  and  $d_2$  if  $y \in \lambda_2$ . If  $f(y/w_i)$  is the conditional frequency function for the signal,  $y$ , then the expected risk will have the following form:

$$E_R = - \sum_{i=1}^2 \sum_{j=1}^2 \int_Y f(y/w_i)p(d_j/y)p(w_i) \log p(w_i/d_j) dy \quad (3)$$

The object is to find the conditions associated with the positioning of the boundary between the two fields which will result in an extremum in the expected risk.

Assume that this extremum exists at some signal value  $y'$  and that  $\lambda_1$  includes all signals between  $y_0$  and  $y'$  while  $\lambda_2$  includes all signals between  $y'$  and  $y_{00}$ . An example of this would be the transmission of a voltage with values between  $y_0$  and  $y_{00}$  which is either signal or noise with the signal represented by a skewed frequency function biased towards the high voltages and the noise represented by a skewed frequency function biased towards the low values. The value for  $y'$  would be the voltage which divides the range into two parts such that  $E_R$  is an extreme. The log term in equation (3) can be rewritten as,

$$\log p(w_i/d_j) = \log \frac{\int_Y f(y/w_i)p(d_j/y)p(w_i)dy}{\int_Y p(d_j/y)f(y)dy}$$

Then,

$$\begin{aligned}
 E_R = & \int_{y_0}^{y'} f(y/w_1) p(w_1) dy \left[ \log \int_{y_0}^{y'} f(y/w_1) p(w_1) dy - \log \int_{y_0}^{y'} f(y) dy \right] \\
 & + \int_{y_0}^{y'} f(y/w_2) p(w_2) dy \left[ \log \int_{y_0}^{y'} f(y/w_2) p(w_2) dy - \log \int_{y_0}^{y'} f(y) dy \right] \\
 & + \int_{y'}^{y_{\infty}} f(y/w_1) p(w_1) dy \left[ \log \int_{y'}^{y_{\infty}} f(y/w_1) p(w_1) dy - \log \int_{y'}^{y_{\infty}} f(y) dy \right] \\
 & + \int_{y'}^{y_{\infty}} f(y/w_2) p(w_2) dy \left[ \log \int_{y'}^{y_{\infty}} f(y/w_2) p(w_2) dy - \log \int_{y'}^{y_{\infty}} f(y) dy \right]
 \end{aligned}$$

At the extreme,

$$\frac{\partial E_R}{\partial y'} = 0 \quad (4)$$

Then,

$$\begin{aligned}
 \frac{\partial E_R}{\partial y'} = & f(y'/w_1) p(w_1) \left[ \log p(w_1, d_1) - \log p(d_1) \right] \\
 & + \int_{y_0}^{y'} f(y/w_1) p(w_1) dy \left[ \frac{f(y'/w_1) p(w_1)}{\int_{y_0}^{y'} f(y/w_1) p(w_1) dy} - \frac{f(y')}{\int_{y_0}^{y'} f(y) dy} \right]
 \end{aligned}$$

EM 1162-141

$$+ f(y'/w_2)p(w_2) \left[ \log p(w_2, d_1) - \log p(d_1) \right]$$

$$+ \int_{y_0}^{y'} f(y/w_2)p(w_2)dy \left[ \frac{f(y'/w_2)p(w_2)}{\int_{y_0}^{y'} f(y/w_2)p(w_2)dy} - \frac{f(y')}{\int_{y_0}^{y'} f(y)dy} \right]$$

$$+ f(y'/w_1)p(w_1) \left[ \log p(w_1, d_2) - \log p(d_2) \right]$$

$$+ \int_{y'}^{y_{\infty}} f(y'/w_1)p(w_1)dy \left[ \frac{f(y'/w_1)p(w_1)}{\int_{y'}^{y_{\infty}} f(y/w_1)p(w_1)dy} - \frac{f(y')}{\int_{y'}^{y_{\infty}} f(y)dy} \right]$$

$$+ \frac{f(y')}{\int_{y'}^{y_{\infty}} f(y)dy} \left[ - f(y'/w_2)p(w_2) \left[ \log p(w_2, d_2) - \log p(d_2) \right] \right]$$

$$+ \int_{y'}^{y_{\infty}} f(y/w_2)p(w_2)dy \left[ - \frac{f(y'/w_2)p(w_2)}{\int_{y'}^{y_{\infty}} f(y/w_2)p(w_2)dy} + \frac{f(y')}{\int_{y'}^{y_{\infty}} f(y)dy} \right]$$

$$\begin{aligned}
\frac{\partial E_R}{\partial y'} &= f(y'/w_1)p(w_1) \frac{\log p(w_1/d_1)}{\log p(w_1/d_2)} - f(y'/w_2)p(w_2) \frac{\log p(w_2/d_2)}{\log p(w_2/d_1)} \\
&\quad + f(y'/w_1)p(w_1) - f(y') \frac{p(w_1, d_1)}{p(d_1)} \\
&\quad + f(y'/w_2)p(w_2) - f(y') \frac{p(w_2, d_1)}{p(d_1)} \\
&\quad - f(y'/w_1)p(w_1) + f(y') \frac{p(w_1, d_2)}{p(d_2)} \\
&\quad - f(y'/w_2)p(w_2) + f(y') \frac{p(w_2, d_2)}{p(d_2)}
\end{aligned}$$

From equation (4),

$$\begin{aligned}
0 &= f(y'/w_1)p(w_1) \frac{\log p(w_1/d_1)}{\log p(w_1/d_2)} - f(y'/w_2)p(w_2) \frac{\log p(w_2/d_2)}{\log p(w_2/d_1)} \\
&\quad - f(y') \frac{p(w_1, d_1) + p(w_2, d_1)}{p(d_1)} + f(y') \frac{p(w_1, d_2) + p(w_2, d_2)}{p(d_2)}
\end{aligned}$$

Thus,

$$\frac{f(y'/w_1)p(w_1)}{f(y'/w_2)p(w_2)} = \frac{\log \frac{p(w_2/d_2)}{p(w_2/d_1)}}{\log \frac{p(w_1/d_1)}{p(w_1/d_2)}}$$

This equation is the basis of a likelihood-ratio test similar to the one described in equation (2) but with a more flexible cost function.

## II. STATISTICAL MODEL

## 1. MODEL WITHOUT LEARNING

For this discussion the signal or pattern considered will have a binary structure. The signal would appear as a series of pulses with amplitude of plus or minus one ( $\pm 1$ ), and the pattern would appear as a cellular matrix with the cells interior to the pattern having a value of +1 and those in the background field having a value of -1. If the set of perfect signals associated with the signal space  $\Omega$  is

$S = [S_1, S_2, \dots, S_n]$  then the set of transmitted signals and the received signal can be considered to be a perturbation of the perfect signal by some noise,  $N$ . Thus, the expression for the signals will be,

$$X_i = S_i N_i = [S_i(1)N_i(1), S_i(2)N_i(2), \dots, S_i(v)N_i(v)]$$

and

$$Y = SN = [Y(1)N(1), Y(2)N(2), \dots, Y(v)N(v)].$$

where  $N_i(c)$  takes on the values of +1 or -1 with the following probability distribution.

$$P [N_i(c) = +1] = p_i(c)$$

$$P [N_i(c) = -1] = 1 - p_i(c) = q_i(c)$$

If the noise components,  $N_i(c)$ , are independent for all  $c$  then the probability distribution for reception of  $y$  given that a member of  $w_i$  was transmitted is the product of the probability distributions for the noise.

$$p(y/w_i) = \prod_{c=1}^v \left[ p_i(c) \right]^{\frac{1+n_i(c)}{2}} \left[ q_i(c) \right]^{\frac{1-n_i(c)}{2}}$$

However,

$$y(c) = s_i(c) \cdot n_i(c)$$

when  $w_i$  is transmitted. Thus,

$$y(c) \cdot s_i(c) = n_i(c)$$

and

$$p(y/w_i) = \prod_{c=1}^v \left[ p_i(c) \right]^{\frac{1+y(c)s_i(c)}{2}} \left[ q_i(c) \right]^{\frac{1-y(c) \cdot s_i(c)}{2}}$$

If the noise components,  $N_i(c)$ , are also identically distributed for all  $c$  then,

$$p(y/w_i) = \left[ p_i \right]^{\frac{v + \sum_{c=1}^v y(c) \cdot s_i(c)}{2}} \left[ q_i \right]^{\frac{v - \sum_{c=1}^v y(c) \cdot s_i(c)}{2}}$$

If the decision rule of equation (1) is applied here, then,

Choose  $d_i$  when:

$$p(w_i) \left[ p_i \right]^{\frac{v + \sum_{c=1}^v y(c) \cdot s_i(c)}{2}} \left[ q_i \right]^{\frac{v - \sum_{c=1}^v y(c) \cdot s_i(c)}{2}} \geq$$

$$p(w_j) \left[ p_j \right]^{\frac{v + \sum_{c=1}^v y(c) \cdot s_j(c)}{2}} \left[ q_j \right]^{\frac{v - \sum_{c=1}^v y(c) \cdot s_j(c)}{2}} .$$

If an additional assumption is made that,

$$p_i = p_j = p$$

the condition for the decision becomes,

$$\frac{\left[ p \right] \frac{\sum_{c=1}^v y(c) \cdot s_i(c)}{2} - \frac{\sum_{c=1}^v y(c) \cdot s_j(c)}{2}}{\left[ q \right] \frac{\sum_{c=1}^v y(c) \cdot s_i(c)}{2} - \frac{\sum_{c=1}^v y(c) \cdot s_j(c)}{2}} \geq \frac{p(w_j)}{p(w_i)} .$$

Taking the logarithm gives,

$$1/2 \left[ \sum_{c=1}^v y(c) \cdot s_i(c) - \sum_{c=1}^v y(c) \cdot s_j(c) \right] \log \frac{p}{q} \geq \log \frac{p(w_j)}{p(w_i)} .$$

If all classes have equal probability and  $p > 1/2$ , the condition for a decision is reduced further to,

$$\sum_{c=1}^v y(c) \cdot s_i(c) \geq \sum_{c=1}^v y(c) \cdot s_j(c)$$

Thus,  $y$  would be classified with the signal which contains the largest number of pulses similar to  $y$  or with the pattern which contains the largest number of cells similar to  $y$ .

## 2. MODEL WITH LEARNING

In many cases of interest the perfect signal is unknown and the noise probability is also unknown. It is desirable then to gain some knowledge about these by incorporating a learning process prior to making a decision. Let the signal  $X_{ik} = [X_{ik}(1), X_{ik}(2), \dots, X_{ik}(v)]$  be the  $k^{\text{th}}$  signal from the class  $w_i$  used in the learning process and

let  $Z_i = [X_{i1}, X_{i2}, \dots, X_{ik_i}]$  be the set of  $k_i$  signals from the  $i^{\text{th}}$  class applied to the learning process. If  $Z$  is the total set of signals for the learning process,

$$Z = [Z_1, Z_2, \dots, Z_n]$$

then the expected risk can be expressed as:

$$E_R = \sum_{Y, Z} \sum_{i=1}^n \sum_{j=1}^n W(w_j/w_i) p(d_j/y) p(y/z, w_i) p(w_i/z) p(z)$$

If the signal classes are independent of the learning observations then

$$p(w_i/z) = p(w_i)$$

Following the reasoning used previously in the model without learning the decision rule becomes,

**DECISION RULE:** Choose class  $k$  with probability 1 and all other classes with probability 0 such that

$$p(d_k/y) = 1$$

and

$$p(d_j/y) = 0 \text{ for } j \neq k$$

when

$$p(w_k)p(y/z_k, w_k) \geq p(w_j)p(y/z_j, w_j) \text{ for } j \neq k \quad (5)$$

In order to evaluate the conditions for this rule the probability distribution  $p(y/z_i, w_i)$  must be computed. This distribution can be represented by the ratio,

$$p(y/z_i, w_i) = \frac{p(y, z_i, w_i)}{p(z_i, w_i)} = \frac{p(y, z_i/w_i)}{p(z_i/w_i)}.$$

EM 1162-141

If the patterns of each class are independent then

$$p(y, z_i/w_i) = p(y/w_i)p(x_{i1}/w_i)p(x_{i2}/w_i) \dots p(x_{ik_i}/w_i)$$

However, from the previous development,

$$p(y/w_i) = \prod_{c=1}^v \left[ p_i(c) \right]^{\frac{1+y(c)s_i(c)}{2}} \left[ q_i(c) \right]^{\frac{1-y(c)s_i(c)}{2}}$$

and,

$$p(X_{ik}/w_i) = \prod_{c=1}^v \left[ p_i(c) \right]^{\frac{1+X_{ik}(c)s_i(c)}{2}} \left[ q_i(c) \right]^{\frac{1-X_{ik}(c)s_i(c)}{2}}$$

Then,

$$p(y, z_i/w_i) = \prod_{c=1}^v \left[ p_i(c) \right]^{\frac{(k_i+1) + s_i(c)}{2} \left[ y(c) + \sum_{k=1}^{k_i} X_{ik}(c) \right]} \cdot \left[ q_i(c) \right]^{\frac{(k_i+1) - s_i(c)}{2} \left[ y(c) + \sum_{k=1}^{k_i} X_{ik}(c) \right]}$$

Since the  $X_{ik}(c)$ 's and the  $y(c)$  are random variables corresponding to  $k_i+1$  observations and the  $p_i(c)$  is a parameter to be estimated, the probability distribution  $p[y(c), z_i(c)/w_i]$  has the form of the likelihood function  $L[y(c), X_{i1}(c), X_{i2}(c), \dots, X_{ik_i}(c)]$ . It is desired to find the

maximum likelihood estimator,  $p_i(c)$ , for  $p_i(c)$ . This can be obtained by differentiating with respect to  $p_i(c)$  and setting the derivative equal to zero. Let  $\sum_{k=1}^{K_i} X_{ik} = U_i$ .

Then,

$$\begin{aligned}
 p \left[ \frac{\partial \log \pi(y(c), z_i(c)/w_i)}{\partial p_i(c)} \right] &= \left\{ p_i(c) \left[ 1 - p_i(c) \right] \right\}^{\frac{K_i+1}{2}} \left\{ \frac{p_i(c)}{1 - p_i(c)} \right\}^{\frac{s_i(c) [u_i(c) + y(c)]}{2}} \\
 \frac{\partial \left[ p \log \pi(y(c), z_i(c)/w_i) \right]}{\partial p_i(c)} &= \left[ \frac{K_i+1}{2} \right] \left[ \frac{\hat{p}_i(c) - \hat{p}_i^2(c)}{\hat{p}_i(c)} \right] \left[ 1 - 2\hat{p}_i(c) \right] \left[ \frac{\hat{p}_i(c)}{1 - \hat{p}_i(c)} \right] \\
 &\quad + \frac{s_i(c) [u_i(c) + y(c)]}{2} \left[ \frac{\hat{p}_i(c)}{1 - \hat{p}_i(c)} \right] - 1 \left[ \frac{s_i(c) [u_i(c) + y(c)]}{2} \right] \\
 &\quad \cdot \left[ \frac{\hat{p}_i(c) - \hat{p}_i^2(c)}{\hat{p}_i(c)} \right]^{\frac{K_i+1}{2}}
 \end{aligned}$$

$$0 = \left[ \frac{k_i+1}{2} \right] \left[ 1-2\hat{p}_i(c) \right] + \frac{s_i(c) [u_i(c)+y(c)]}{2} \left[ \frac{1-\hat{p}_i(c)}{\hat{p}_i(c)} \right] \left[ \frac{1}{[1-\hat{p}_i(c)]^2} \right].$$

$$\left[ \hat{p}_i(c) \quad [1-\hat{p}_i(c)] \right]$$

$$0 = \left[ \frac{k_i+1}{2} \right] \left[ 1-2\hat{p}_i(c) \right] + \frac{s_i(c) [u_i(c)+y(c)]}{2}.$$

Thus the estimator is,

$$\hat{p}_i(c) = \frac{(k_i+1)+s_i(c) [u_i(c)+y(c)]}{2(k_i+1)}$$

The estimator for  $p_i(c)$  in the probability distribution  $p(y_i/w_i)$  is

$$\hat{p}_i(c) = \frac{k_i+s_i(c)u_i(c)}{2k_i}$$

For large  $k_i$ ,

$$\frac{\hat{p}_i(c)}{\hat{p}_i(c)} = \frac{\frac{(k_i+1)+s_i(c) [u_i(c)+y(c)]}{2(k_i+1)}}{\frac{k_i+s_i(c)u_i(c)}{2k_i}} = \frac{\frac{k_i+s_i(c)u_i(c)}{k_i+1} + \frac{1+s_i(c)y(c)}{k_i+1}}{\frac{k_i+s_i(c)u_i(c)}{k_i}}$$

$$\frac{\hat{p}_i(c)}{\hat{p}_i(c)} \approx 1.0$$

Then, the probability distribution  $p(y/z_i, w_i)$  becomes

$$p(y/z_i, w_i) = \frac{\prod_{c=1}^v \left[ \hat{p}_i(c) \right]^{\frac{(k_i+1)+s_i(c) [y(c)+u_i(c)]}{2}} \left[ \hat{q}_i(c) \right]^{\frac{(k_i+1)-s_i(c) [y(c)+u_i(c)]}{2}}}{\prod_{c=1}^v \left[ \hat{p}_i(c) \right]^{\frac{k_i+s_i(c)u_i(c)}{2}} \left[ \hat{q}_i(c) \right]^{\frac{k_i-s_i(c)u_i(c)}{2}}}$$

or,

$$p(y/z_i, w_i) = \prod_{c=1}^v \left[ \hat{p}_i(c) \right]^{\frac{1+s_i(c)y(c)}{2}} \left[ \hat{q}_i(c) \right]^{\frac{1-s_i(c)y(c)}{2}},$$

If it is assumed that the learning process is biased toward the perfect signal such that  $p_i(c) > 1/2$ ,

$$s_i(c) = \text{sign } u_i(c) = \text{sign } \sum_{k=1}^{k_i} x_{ik}(c).$$

With this assumption the probability distribution becomes,

$$p(y/z_i, w_i) = \prod_{c=1}^v \left[ \hat{p}_i(c) \right]^{\frac{1+y(c) \cdot \text{sign } \sum_{k=1}^{k_i} x_{ik}}{2}} \left[ \hat{q}_i(c) \right]^{\frac{1-y(c) \text{sign } \sum_{k=1}^{k_i} x_{ik}}{2}} \quad (6)$$

where

$$\hat{p}_i(c) = \frac{(k_i+1) + \text{sign} \sum_{k=1}^{k_i} x_{ik}(c) \left[ y(c) + \sum_{k=1}^{k_i} x_{ik}(c) \right]}{2(k_i+1)} \quad (7)$$

It would facilitate the mechanization of the decision if the rule expressed in equation (1) could be put into the linear form,

$$G_k + \sum_{c=1}^n F_k(c)y(c) \geq G_j + \sum_{c=1}^v F_j(c)y(c)$$

This can be accomplished by letting

$$G_k + \sum_{c=1}^v F_k(c)y(c) = \log p(w_k)p(y/z_k, w_k).$$

Then,

$$G_k + \sum_{c=1}^v F_k(c)y(c) = \log p(w_k) + \log p(y/z_k, w_k)$$

or,

$$G_k + \sum_{c=1}^v F_k(c)y(c) = \log p(w_k) + \sum_{c=1}^v \left[ \frac{1+y(c)\text{sign } u_k(c)}{2} \log \hat{p}_k(c) + \frac{1-y(c)\text{sign } u_k(c)}{2} \log \hat{q}_k(c) \right]$$

$$G_k + \sum_{c=1}^v F_k(c)y(c) = \log p(w_k) + \sum_{c=1}^v \frac{1}{2} \left[ \log \hat{p}_k(c) + \log \hat{q}_k(c) \right] + \sum_{c=1}^v \frac{1}{2} y(c) \text{sign } u_k(c) \cdot \left[ \log \hat{p}_k(c) - \log \hat{q}_k(c) \right] \quad (8)$$

Thus,

$$G_k = \log p(w_k) + \sum_{c=1}^V \frac{1}{2} \left[ \log \hat{p}_k(c) + \log \hat{q}_k(c) \right]$$

and

$$F_k(c) = \frac{1}{2} \text{sign } u_k(c) \left[ \log \hat{p}_k(c) - \log \hat{q}_k(c) \right]$$

### 3. INFORMATION THEORY

If the value for  $y(c)$  in equation (8) has the same sign as  $u_k(c)$  then,

$$G_k + F_k(c) \cdot y(c) = \log p(w_k) + \log \hat{p}_k(c).$$

If  $y(c)$  possesses the opposite sign of  $u_k(c)$  then,

$$G_k + F_k(c)y(c) = \log p(w_k) + \log \hat{q}_k(c)$$

In "information theory" the terms

$$I(w_k) = -\log p(w_k)$$

and

$$I(c/w_k) = -\log \hat{p}_k(c)$$

or

$$I(c/w_k) = -\log \hat{q}_k(c)$$

are referred to as the "self-information." Using the definitions of "information theory," the decision rule would be a summation of the self-information associated with a particular class plus the self-information associated with the learning process for each component or characteristic of that class. If the probability of the classes are equal i.e.,  $p(w_k) = p(w_j)$ , the decision rule is based upon a

comparison of the self-information of a particular component or characteristic of one class against the self-information of this same characteristic in another class. The average, or expected amount of self-information for a specified component of one class, is expressed in equation (9) where the following nomenclature change has been made to be consistent with other reports on information theory:

$$p_k(c) = p(c/w_k) \text{ and } q_k(c) = q(c/w_k)$$

$$E \left[ I(c/w_k) \right] = p(c/w_k) \log p(c/w_k) + q(c/w_k) \log q(c/w_k). \quad (9)$$

The average self-information for one characteristic and for all classes then is,

$$E \left[ I(c/\Omega) \right] = \sum_{k=1}^n p(w_k) \left[ p(c/w_k) \log p(c/w_k) + q(c/w_k) \log q(c/w_k) \right]. \quad (10)$$

A question of information rate arises when an attempt is made to evaluate the information obtained about a particular class of signals by comparing the self-information of a specified component for two classes. That is, which components provide significant information about the transmitted signal class and which ones could be ignored or not even transmitted without loss of information about the signal class? A measure of the importance is contained in the statistical dependence between the characteristics and the signal. Information theory suggests the following function for measuring this quality.

$$I(c, w_k) = \log \frac{p(c, w_k)}{p(c)p(w_k)}$$

or,

$$I(c, w_k) = \log \frac{q(c, w_k)}{q(c)p(w_k)}$$

This term is referred to as "mutual-information" and the expected or average mutual information is,

$$E \left[ I(c, w_k) \right] = \sum_{k=1}^n p(c, w_k) \log \frac{p(c, w_k)}{p(c)p(w_k)} + q(c, w_k) \log \frac{q(c, w_k)}{q(c)p(w_k)}.$$

Expanding, this gives,

$$\begin{aligned}
 E \left[ I(c, w_k) \right] &= \left[ \sum_{k=1}^n p(w_k) p(c/w_k) \cdot \log p(c/w_k) + p(w_k) q(c/w_k) \cdot \log q(c/w_k) \right] \\
 &\quad - \sum_{k=1}^n p(c, w_k) \log p(c) + q(c, w_k) \log q(c) \\
 E \left[ I(c, w_k) \right] &= - \left\{ p(c) \log p(c) + q(c) \log q(c) - \sum_{k=1}^n p(w_k) \left[ p(c/w_k) \right. \right. \\
 &\quad \left. \left. \cdot \log p(c/w_k) + q(c/w_k) \log q(c/w_k) \right] \right\} \quad (11)
 \end{aligned}$$

Thus the expected mutual-information, referred to as "transinformation," is the expected self-information in the component, unconditioned by the signals, minus the expected self-information of equation (10). These terms are referred to as entropy and designated by  $H$ . Using the nomenclature of information theory, equation (11) becomes.

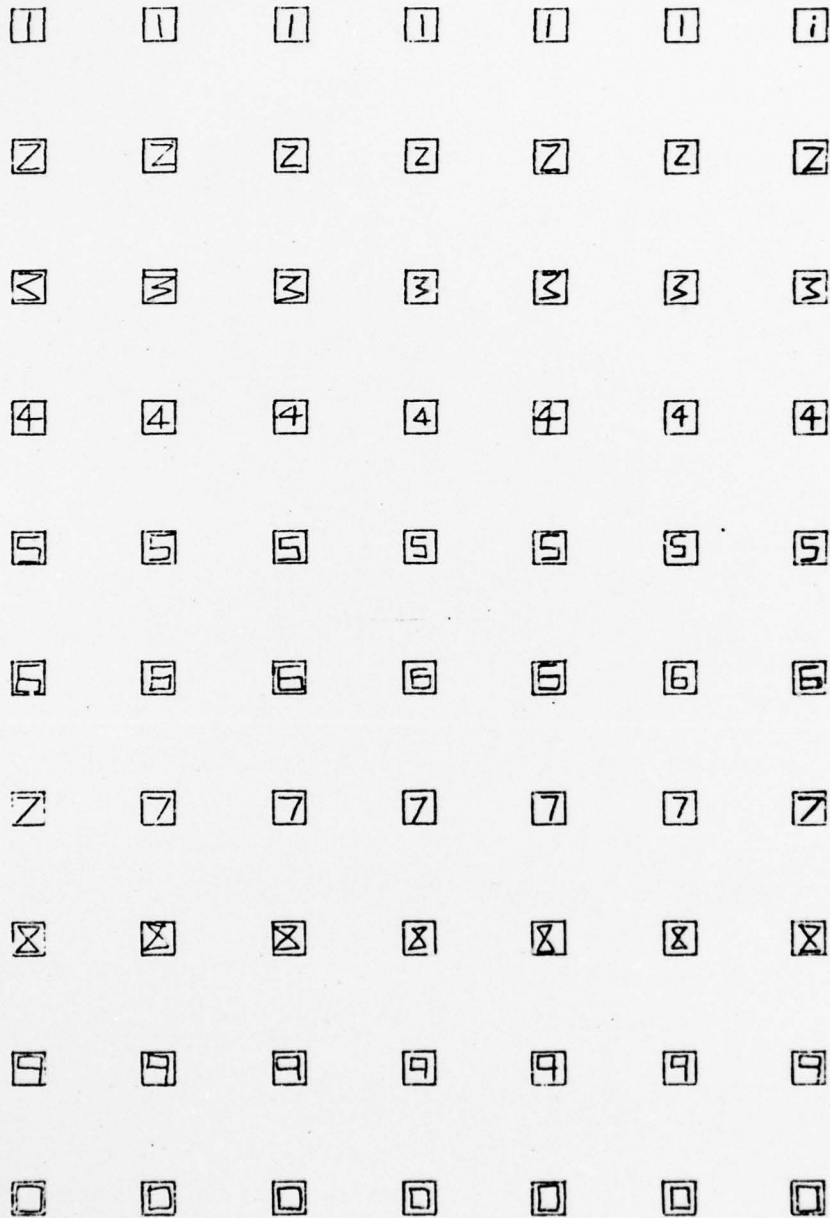
$$I(C, \Omega) = H(C) - H(C/\Omega)$$

#### 4. AN EXPERIMENT WITH WRITTEN NUMERALS

In order to explore the applications of the theory developed above, a simple experiment was performed with seven samples of the ten numbers 0 through 9. The experimental data were collected by having seven individuals print the numbers 0 through 9 inside a 1/4-inch square (see Figure 1). The formation of these numbers was controlled to the extent that an original pattern set was provided as an example and only straight lines were permitted. The numbers were then magnified and quantized into a 16 x 16 cellular matrix as shown for the number 4 in Figure 2. The first six samples were chosen as the learning set and the seventh sample was used to check the effectiveness of the recognition scheme.

Three sets of data were tabulated from the learning patterns for each pattern class. The first set of data was  $\sum_{k=1}^K X_{ik}$  for each cell of

EM 1162-141



Learning Set

Figure 1. Printed Numerals Used in Experiment

EM 1162-141

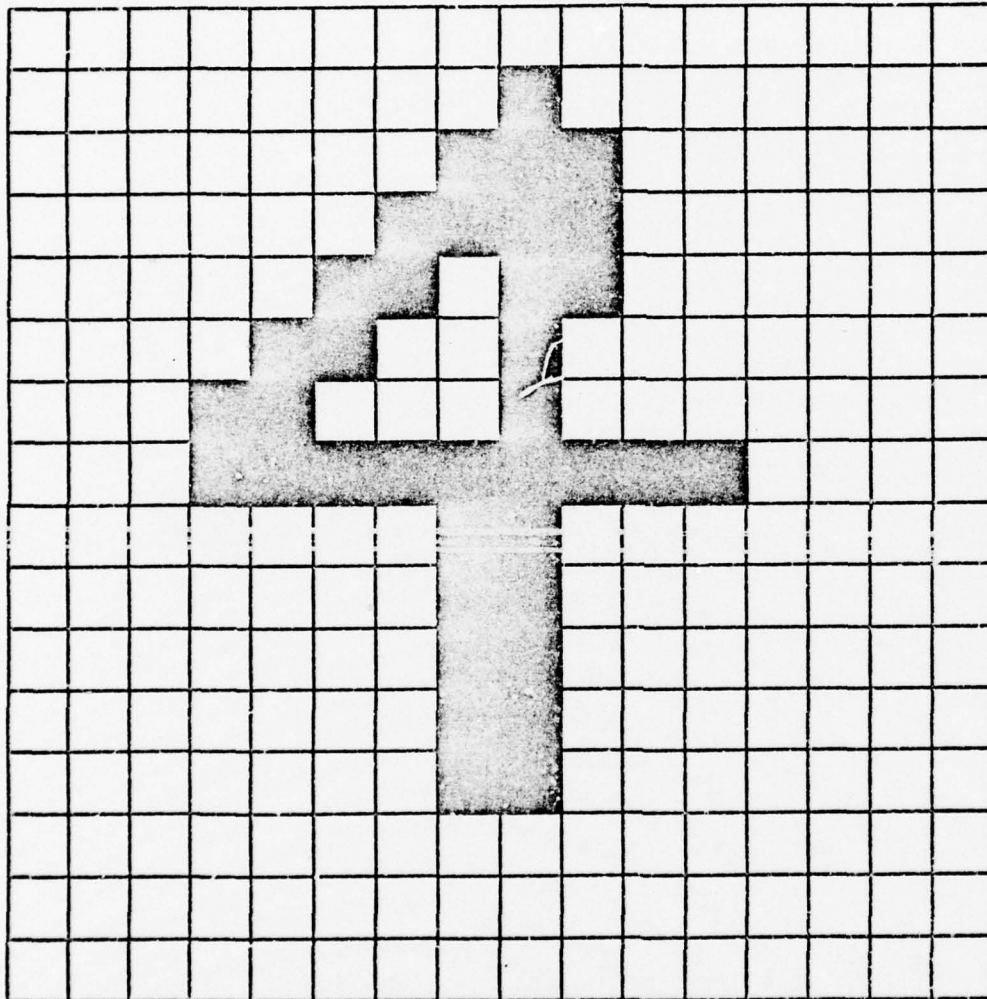


Figure 2. The Number 4 Quantized Into a 16 x 16 Cellular Matrix.  
This is the Number 4 Which Appears in the Last Column  
of the Hand Printed Numerals.

the 16 by 16 matrix. An example of the resulting values for the number 4 is shown in Figure 3. The other two sets of data involved the generation of a transformed pattern from the original by observing the number of pattern crossings a scanning system would encounter in the vertical and horizontal directions if the scanning resolution was comparable to the quantizing of the pattern into the 16 x 16 matrix. The transformed patterns for the vertical scan are embodied in a 4 x 16 cellular matrix in which the 16 columns represent the 16 scan lines and the 4 rows represent the number of crossings encountered. An example of this matrix for a 4 is shown in Figure 4. The transformed patterns obtained by horizontal crossing are contained in a 2 x 16 matrix and an example of this for the number 4 is shown in Figure 5. Each cell of the 4 x 16 matrix and the 2 x 16 matrix was then treated in the same manner as the cells of the original pattern

matrix in order to establish the statistic  $\sum_{k=1}^{k_i} X_{ik}$ . Examples of these are shown in Figures 6 and 7.

The decision rule of equation (5) and the statistical model developed for equations (6) and (7) were applied to the statistics of the transformed patterns as a recognition scheme for the seventh sample of patterns. The form of the decision rule which was used here was the summation of the logarithm of the probabilities. The results for the horizontal crossings are shown in Table 1. The encircled values denote the pattern classes where confusion has occurred because the values are larger than those for the proper pattern classes which occur along the diagonal. Table 2 represents the same recognition scheme for the vertical crossings. The misrecognition present in these two transformed patterns can be eliminated by combining the two as a single transformed pattern in a 6 by 16 matrix. The results of the recognition scheme for the combined transformed patterns is shown in Table 3.

A simplification of the decision rule and the statistical model was applied to the statistics of the 16 x 16 matrix of the original patterns for recognition of the seventh sample of patterns. This simplification involved the summing of the probabilities for the cells instead of the logarithm of the probabilities where the probabilities were computed from the following equation,

$$\hat{p}_i(c) = \frac{k_i + \text{sign} \sum_{k=1}^{k_i} X_{ik}(c) \left[ \sum_{k=1}^{k_i} X_{ik}(c) \right]}{2k_i} .$$

EM 1162-141

-6	-6	-6	-6	-6	-6	-6	-2	-2	-6	-6	-6	-6	-6	-6
-6	-6	-6	-6	-6	-6	-6	-2	0	-4	-6	-6	-6	-6	-6
-6	-6	-6	-6	-6	-6	-2	0	-2	+2	-4	-6	-6	-6	-6
-6	-6	-6	-6	-6	-2	0	0	+2	0	-4	-6	-6	-6	-6
-6	-6	-6	-6	-2	+2	-2	+2	-2	0	-4	-6	-6	-6	-6
-6	-6	-6	-4	0	-2	-6	0	0	0	-4	-6	-6	-6	-6
-6	-6	-2	0	-4	-4	-4	0	0	+2	-2	-6	-6	-4	-6
-6	-2	0	0	-2	0	-2	+2	0	+2	-2	-4	-4	-4	-6
-4	-2	+2	+4	+4	+2	0	0	+2	+2	+4	0	0	0	-4
-6	-4	-4	0	-2	-4	-2	-2	-2	+2	-2	-6	-6	-6	-6
-6	-6	-4	-4	-2	-2	-2	0	+2	+2	0	-2	-2	-4	-6
-6	-6	-4	-4	-4	-4	-4	-2	0	+4	-2	-6	-6	-6	-6
-6	-6	-6	-6	-6	-6	-6	-2	-2	+2	-4	-6	-6	-6	-6
-6	-6	-6	-6	-6	-6	-6	-2	-2	0	-4	-6	-6	-6	-6
-6	-6	-6	-6	-6	-6	-4	-2	-2	-4	-6	-6	-6	-6	-6
-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6

Figure 3. A Compilation of the Values for  $\sum_{k=1}^{k_i} X_{ik}$  for the Hand  
Printed 4s Occurring in the First Six Columns in Figure 1.  
(Specifically These are Values for  $\sum_{k=1}^6 X_{4k}$ )

EM 1162-141

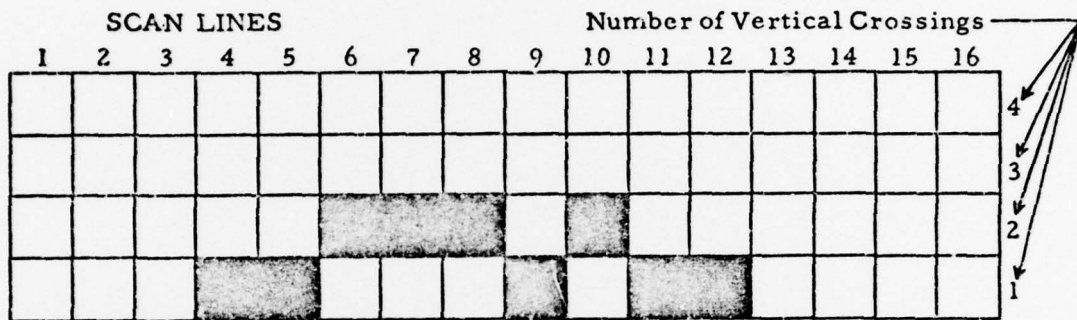


Figure 4. Transformed Pattern for the Vertical Scan Lines of the Number 4 Shown in Figure 2.

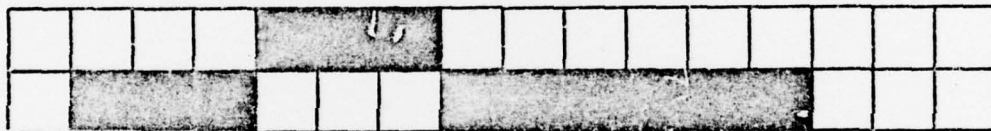


Figure 5. Transformed Pattern for the Horizontal Scan Lines of the Number 4 Shown in Figure 2.

-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6
-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6
-6	-6	-6	-4	+2	+4	+6	+2	+2	-2	-6	-6	-6	-6	-6	-6
-2	-2	+2	+4	-2	-4	-6	-2	-2	+4	+6	+6	+6	+4	-2	-6

Figure 6.  $\sum_{k=1}^{k_1} X_{ik}$  For the Vertical Scan Pattern of the 4s.

-6	-6	-6	-6	+4	+4	+6	+6	-2	-4	-6	-6	-6	-6	-6	-6
-6	+2	+4	+4	-4	-4	-6	-6	+2	+4	+6	+6	+6	+2	0	-6

Figure 7.  $\sum_{k=1}^{k_1} X_{ik}$  For the Horizontal Scan Patterns of the 4s.

Table 1. Recognition Matrix for the Seventh Set of Patterns Formed  
by Summing the Log of the Probabilities for the 64 Cells of the  
Horizontal Crossing Transformation

$$\sum_{k=1}^{64} \log p_k \left[ y(c)/w_i \right]$$

7th Sample	0	1	2	3	4	5	6	7	8	9
0	-9.91	-25.01	-23.53	-24.87	-17.50	-32.20	-35.63	-8.56	-29.70	-9.29
1	-24.49	-1.00	-18.20	-18.00	-18.05	-20.84	-23.11	-15.92	-18.17	-21.75
2	-16.64	-20.16	-14.10	-20.82	-25.64	-19.06	-24.39	-15.15	-18.34	-16.58
3	-28.23	-22.11	-19.10	-17.70	-29.84	-19.05	-21.68	-23.11	-16.09	-25.24
4	-20.59	-17.11	-23.78	-29.95	-11.91	-33.69	-32.29	-15.05	-30.13	-15.90
5	-26.90	-20.16	-15.12	-16.64	-34.15	-10.50	-17.27	-29.66	-12.98	-25.43
6	-32.29	-20.16	-11.26	-17.82	-31.95	-9.39	-9.47	-25.99	-14.27	-31.38
7	-21.10	-20.16	-27.82	-30.80	-10.20	-34.49	-35.38	-10.47	-30.97	-15.90
8	-12.79	-20.16	-16.77	-14.52	-31.74	-16.60	-23.63	-20.14	-19.44	-20.48
9	-13.30	-22.11	-25.53	-29.95	-12.84	-34.15	-34.82	-9.55	-30.63	-9.80

Table 2. Recognition Matrix for the Seventh Set of Patterns Formed  
by Summing the Log of the Probabilities for the 32 Cells of the  
Vertical Crossing Transformation.

$$\sum_{k=1}^{32} \log P_k \left[ y(c)/w_i \right]$$

7th Sample	$w_i$	0	1	2	3	4	5	6	7	8	9
0		-8.22	-33.22	-33.26	-34.12	-14.73	-33.12	-18.48	-32.18	-17.28	-24.19
1		-23.82	-2.02	-2.06	-3.14	-15.97	-2.08	-19.78	-2.93	-28.67	-18.34
2		-38.32	-5.92	-4.52	-7.03	-18.38	-4.97	-18.68	-6.67	-26.87	-23.18
3		-37.20	-5.25	-5.64	-4.24	-16.85	-4.72	-21.06	-3.18	-24.07	-17.41
4		-27.21	-16.77	-15.37	-17.29	-7.61	-15.73	-27.24	-17.68	-21.36	-9.43
5		-38.06	-6.51	-5.28	-7.89	-22.96	-7.59	-20.05	-9.03	-28.48	-17.16
6		-23.31	-18.72	-18.76	-19.24	-21.49	-18.19	-7.74	-19.63	-19.58	-30.88
7		-40.42	-3.97	-2.57	-5.08	-16.42	-3.18	-19.28	-4.88	-27.12	-22.23
8		-23.82	-21.52	-21.56	-22.58	-22.36	-21.58	-23.68	-20.73	-10.17	-17.92
9		-27.67	-14.31	-15.98	-13.39	-14.05	-13.19	-30.46	-12.08	-17.97	-9.69

Table 3. Recognition Matrix for the Combined Results of  
Tables 1 and 2.

7th Sample $w_i$	0	1	2	3	4	5	6	7	8	9
0	-19.13	-58.23	-56.79	-58.99	-31.23	-65.32	-54.11	-40.74	-45.98	-33.48
1	-48.31	-3.02	-20.26	-21.14	-34.02	-22.92	-42.89	-18.85	-46.84	-40.09
2	-54.96	-26.08	-18.62	-27.85	-44.02	-24.03	-43.07	-21.82	-45.21	-39.76
3	-65.43	-27.36	-24.74	-21.94	-46.69	-23.75	-42.74	-26.29	-40.16	-42.65
4	-40.80	-33.88	-39.15	-47.24	-19.52	-49.12	-59.53	-32.73	-51.49	-25.33
5	-64.96	-26.67	-20.40	-24.53	-57.11	-18.09	-31.32	-38.69	-41.46	-42.69
6	-55.60	-38.88	-30.02	-37.06	-53.44	-27.58	-17.21	-45.62	-33.85	-52.26
7	-61.52	-24.13	-30.39	-35.88	-26.62	-37.67	-54.66	-15.35	-58.09	-38.13
8	-36.61	-41.68	-38.33	-37.10	-54.10	-38.18	-47.31	-40.87	-29.61	-38.40
9	-40.97	-36.42	-41.51	-43.34	-46.19	-47.34	-65.28	-21.63	-48.60	-19.49

EM 1162-141

The results of this recognition scheme are shown in Table 4. Once again the enclosed values represent pattern classes where confusion has occurred. Only three of the ten numerals are correctly classified for this particular scheme.

It would seem, from the discussion of information theory above, that a large number of cells could be eliminated from the  $16 \times 16$  matrix on the basis of transinformation content without incurring a proportionate deterioration in recognition capabilities. In order to evaluate this scheme, the transinformation for each of the 256 cells was computed. These values are shown in Figure 8. A transinformation value of 0.30 was arbitrarily chosen as a lower limit and all cells with values equal to or greater than this were used with the simplified recognition scheme described above to recompute the recognition matrix. The distribution of the 55 cells (approximately  $1/5$  of the total number) with values equal to or greater than 0.30 is shown in Figure 9. The results of the recognition scheme are shown in Table 5. A comparison between these results and the values obtained by using all of the cells, as presented in Table 6, shows a reduction in the number of confusions from 20 to 10 and an increase in the number of correct recognitions from 3 to 5. As a contrast, the 55 cells with the lowest transinformation content (See Figure 10) were also used with this recognition scheme. These results, presented in Table 6, indicate how insignificant their contribution is with regards to recognition of the pattern classes.

EM 1162-141

0	0	.06	.08	.06	.06	0	.12	.13	.05	.08	.08	.06	0	.06	.06
.06	.06	.30	.38	.23	.24	.26	.20	.12	.19	.22	.20	.16	.09	.08	.06
.06	.08	.26	.41	.28	.29	.20	.13	.19	.29	.21	.27	.28	.26	.14	.06
0	.09	.38	.26	.25	.15	.11	.12	.11	.18	.27	.24	.33	.30	.15	.08
0	.08	.30	.41	.15	.27	.12	.30	.23	.33	.26	.26	.21	.21	.13	0
0	.08	.27	.40	.23	.15	.13	.27	.42	.25	.24	.24	.21	.23	.17	0
0	.08	.25	.41	.10	.17	.30	.36	.30	.25	.14	.24	.15	.27	.20	0
0	.11	.27	.45	.34	.33	.30	.34	.29	.36	.43	.53	.41	.21	.21	0
.06	.13	.30	.47	.42	.38	.28	.13	.17	.29	.44	.33	.38	.39	.21	0
0	.08	.24	.37	.23	.32	.27	.35	.21	.24	.15	.14	.21	.32	.27	0
0	.06	.23	.30	.24	.27	.27	.25	.34	.34	.23	.23	.15	.34	.25	0
0	.08	.25	.45	.19	.22	.13	.13	.24	.41	.25	.25	.35	.40	.30	0
0	.13	.34	.40	.28	.13	.11	.14	.14	.18	.14	.32	.23	.25	.25	.08
.12	.17	.34	.29	.29	.29	.24	.23	.22	.23	.27	.31	.31	.24	.18	.06
.06	.17	.20	.33	.35	.24	.18	.13	.13	.21	.25	.17	.21	.15	.13	0
0	.12	.11	.11	.13	.11	.11	.11	.09	.11	.11	.09	.11	.11	0	0

Figure 8. Transinformation Values as Computed From  
Equation 10 for Each of the Cells in the  
16 x 16 Matrix.

EM 1162-141

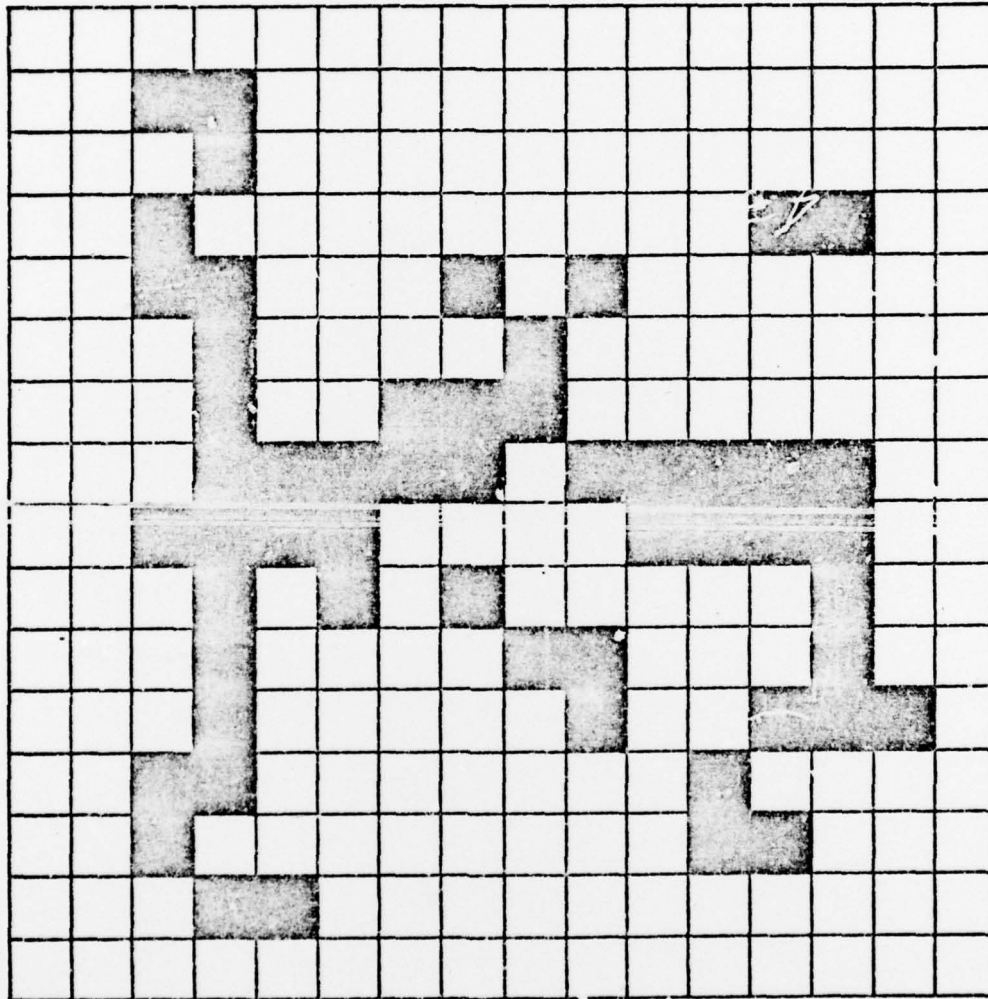


Figure 9. The Blackened Cells Have a Transinformation Value Equal to or Greater Than 0.30  
(See Figure 8).

EM 1162-141

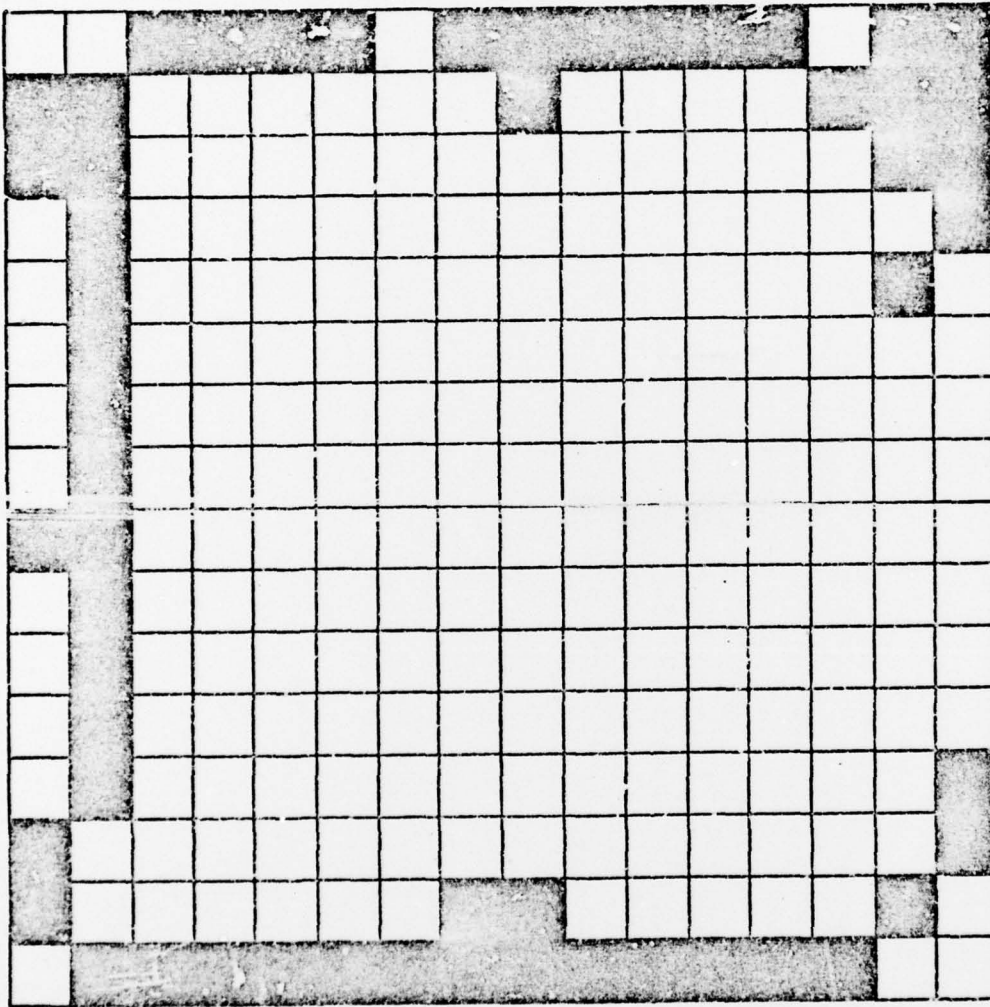


Figure 10. The Blackened Cells Have a Transinformation  
Equal to or Less Than 0.13  
(See Figure 8).

Table 4. Recognition Matrix for the Seventh Set of Patterns Formed  
by Summing the Probabilities for the 256 Cells as Developed  
from the Learning Set of Patterns.

$$\sum_{k=1}^{256} P_k \left[ \frac{y(c)}{w_i} \right]$$

7th Sample \ $w_i$	0	1	2	3	4	5	6	7	8	9
0	190	177	179	171	169	186	185	188	158	179
1	177	239	220	198	211	188	178	205	183	191
2	171	218	205	192	210	185	173	222	185	188
3	169	208	190	192	194	183	171	200	179	175
4	166	224	193	191	209	188	176	199	176	174
5	165	192	179	173	215	180	175	182	176	188
6	171	191	182	174	189	199	189	187	173	184
7	177	223	205	191	204	187	177	216	181	192
8	163	199	191	187	187	174	171	191	186	165
9	164	186	181	173	180	194	167	192	179	198

Table 5. Recognition Matrix for the Seventh Set of Patterns Formed by Summing the Probabilities for the Blackened 55 Cells of Figure 9.

$$\sum_{55 \text{ cells}} P_k \left[ y(c)/w_i \right]$$

7th Sample \ $w_i$	0	1	2	3	4	5	6	7	8	9
0	37	29	31	22	26	34	34	30	24	35
1	26	50	42	39	39	29	24	41	36	30
2	26	46	41	39	35	29	24	40	36	30
3	19	39	37	41	35	26	23	36	36	26
4	19	41	39	35	37	35	26	38	29	35
5	24	33	31	27	30	34	28	33	31	33
6	28	31	33	26	34	37	35	32	29	34
7	32	49	46	37	38	34	28	45	35	34
8	22	43	38	40	34	25	22	36	38	30
9	23	34	34	29	32	39	30	35	31	42

Table 6. Recognition Matrix for the Seventh Set of Patterns Formed  
by Summing the Probability for the Blackened Cells  
Shown in Figure 10.

$$P_k \left[ y(c)w_i \right]$$

55 cells

7th Sample	$w_i$	0	1	2	3	4	5	6	7	8	9
0	48	54	52	52	51	52	50	51	54	47	51
1	48	54	52	52	51	52	50	51	54	47	51
2	48	54	52	52	51	52	50	51	54	47	51
3	48	54	52	52	51	52	50	51	54	47	51
4	47	54	52	52	51	52	50	51	53	47	52
5	47	54	52	52	51	52	50	51	53	47	52
6	48	54	52	52	51	52	50	51	54	47	51
7	48	54	52	52	51	52	50	51	54	47	51
8	48	54	52	52	51	52	50	51	54	47	51
9	47	54	52	52	51	52	50	51	54	47	52

### III. CONCLUSIONS

The inadequacies of the experiment preclude any quantitative conclusions. However, the results do indicate two areas in which further investigations could be rewarding. One is the design of a recognition scheme which would automatically read controlled hand printed characters. The other is the development of a procedure for evaluating and weighting pattern characteristics with respect to their discriminating qualities.

A certain degree of control is required in the printing of characters even when the recognition is to be performed by humans. It is reasonable to expect that the complexity of a machine to recognize hand printed characters is dependent to a large extent upon how much responsibility can be put upon the printer for controlling the distortion of the characters. In the area of the computer programmer, a number of restrictions have already been applied with regards to where and how the program is to be written and the legibility required. If additional controls (which require the programmer to imitate a particular set of straight line numerals inside of a box or a printed form) were acceptable it appears that a simple pattern recognition scheme could be devised to automatically read and punch computer programs. Of course, the problem is two sided in that the amount of distortion acceptable to the recognition scheme must not be too restrictive with respect to the human capabilities. In order to answer questions arising from the problems, further studies would be required to establish machine complexity versus distortion of characters. Further studies would also be required to define the capabilities of humans to print controlled characters, including considerations for the time element and increased distortion caused by fatigue.

The area of evaluation of pattern characteristics has general applications. It has been shown that transinformation has a definite bearing on the discriminating qualities of the characteristic. However, the linearity of this relationship and the effects of statistical dependence between characteristics has not been considered. In order to establish some quantitative measures in this area it would be necessary to perform a more comprehensive experiment on some particular set of pattern classes. The initial problems confronting such an experiment are the compiling of the sampled patterns, data processing for computer application, and computer programming to tabulate the statistics. It would also be of considerable interest to perform a comparison of the evaluation of characteristics obtained with information theory techniques and the adaptive memory process.

EM 1162-141

REFERENCES

- (1) Braverman, D. J. , "Machine Learning and Automatic Pattern Recognition. "
- (2) Middleton, David, "Introduction to Statistical Communication Theory. "
- (3) Reya, F. M. , "An Introduction to Information Theory. "

EM 1162-141

REFERENCES

- (1) Braverman, D. J. , "Machine Learning and Automatic Pattern Recognition. "
- (2) Middleton, David, "Introduction to Statistical Communication Theory. "
- (3) Reya, F. M. , "An Introduction to Information Theory. "